ACTIVITY 5: Worksheet Answers

In this activity, you will be exploring the one sample t-confidence interval as an estimate for the population mean.

The general formula for the confidence interval is:

$$\bar{x} - E < \mu < \bar{x} + E,$$

where $\bar{x}$ represents the sample mean, E represents the marginal error, and μ represents the population mean.

Answer the following questions before calculating the confidence intervals.

1. Since you are subtracting the margin of error from the sample mean to find the lower bound and adding the margin of error to the mean to find the upper bound of the confidence interval, where will the sample mean be located in your confidence interval?

   The sample mean will be the midpoint of the confidence interval.

2. Based on your knowledge from the previous activities, predict how the sample size, impacts the width of the confidence interval.

   The larger the sample size, the narrower the confidence interval.

3. When calculating confidence intervals, you choose a confidence level such as 80%, 90%, 95%, or 99%. As presented in the lesson, the confidence level tells you how confident you are that for the sample you selected, the interval you calculate will capture the population mean you are trying to estimate. Based on this information, predict how the confidence level impacts the width of the confidence interval.

   If you want to be more confident that your interval contains the population mean, meaning you would have a higher confidence level (e.g. 95%), your confidence interval would most likely be wider as a wider interval is more likely to contain the population mean.

You are now ready to calculate some confidence intervals for each of your samples. You will use the sample means and the sample standard deviations you found in Activity 2 for the various sample sizes. Below is a link to a confidence interval calculator. Use this calculator to find confidence intervals for each of the sample sizes and confidence levels.

https://www.socscistatistics.com/confidenceinterval/default2.aspx

Answers will vary. However, for the example data given in Activity 2, the intervals are as follows.

Confidence Level	n=5 Sample	n=15 Sample	n=30 Sample
80%	(67.253, 76.227)	(69.434, 73.946)	(71.132, 74.608)
90%	(65.501, 77.979)	(68.736, 74.644)	(70.618, 75.129)
95%	(63.615, 79.865)	(68.093, 75.287)	(70.159, 75.581)
99%	(58.266, 85.214)	(66.698, 76.682)	(69.217, 76.523)

4. Looking at the sample confidence intervals above, what happens to the width of the interval as the sample size increases for each confidence level?

   As the sample size increases, the confidence interval becomes more narrow for each of the confidence levels. Consequently, you get a more precise estimate.

5. How does the confidence level impact the width of the confidence interval for each sample size group?

   As the level of confidence increases, the confidence interval becomes wider for each sample size group.
   Note to teacher: When you discuss these questions with the students in an all-class discussion, point out the trade-off between accuracy (the likelihood of the confidence interval capturing the parameter) and precision (the variability of the underlying sampling distribution reflected by the width of the confidence interval).

6. How useful would the 99% confidence interval for the n=5 group be as an estimate of the population mean? Explain your answer.
   The confidence interval for the *n*=5 group lacks utility because of how wide it is.

7. Did your 90% confidence interval for the *n*=30 sample size group capture the population mean of $\mu = 73.76$?
   For our group, it did capture the population mean.
   Note to teacher: There may be some groups where the confidence interval does not capture the population mean.

8. As your instructor pointed out, when wording a confidence interval, you would state the following:

   There is ______% confidence that the interval (______, ______) captures the population mean……..

   Using the format above, state your 90% confidence interval for the sample size of n=30.

   Answers will vary. For our example, the statement would read:

   There is 90% confidence that the interval (70.618, 75.129) captures the population mean life expectancy for countries.

9. Based on that wording, do you think it is possible to select a sample and calculate a confidence interval that does not capture the population mean?

   Yes, since you are saying there is a 90% chance of your confidence interval capturing the population mean, there is a 10% chance that you could select a sample where your confidence interval does not capture the population mean.

The confidence interval formula we used is only valid when the following assumptions are met. You will be considering these assumptions in relation to your confidence intervals in the following questions.

Assumptions for One Sample T-Confidence interval:

- The units or subjects need to be selected using a simple random sample.
  If you are using a biased non-random sample, the formulas will yield biased results.
- The distribution of the data needs to be approximately normally distributed. The smaller the sample size, the more normal the data needs to be. Typically, if the sample size is 30 or more, it is fine to use the T-procedure even when the data is not normally distributed or there are non-extreme outliers.

10. What type of sampling design did you use to collect your data in Activity 2?

    We used a simple random sample.

11. Does that satisfy the first assumption?

    Yes, the correct sampling design was used.

You now need to determine whether the second assumption is satisfied.
While we are unable to create a histogram for our population distribution, we can create one for our sample to get an idea of how the population data might be distributed.

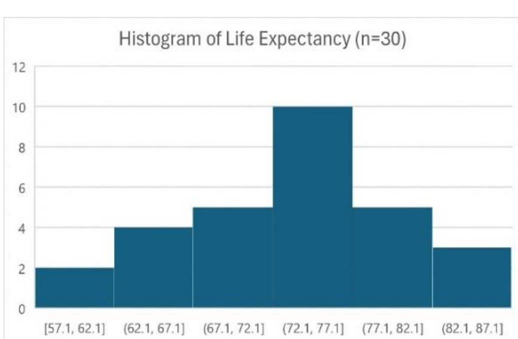
12. Use Excel to create a histogram of the life expectancy data for the 30 countries you randomly selected in Activity 2.
Recall: You should have saved this in a spreadsheet in Activity 3.

Steps for Creating a Histogram

Instructions for a Windows Device	**Instructions for an Apple Device**
1. Use the saved spreadsheet you created for Activity 3 of your data. 2. Highlight the column that has the life expectancies for the sample size of 30. 3. Select "Insert" in the top menu. 4. To get the histogram, click the icon that looks like a histogram (all bars are blue) and select the histogram picture on the left. 5. A histogram should pop up on your data editor. 6. Click on the box that says, "Chart Title" and create an appropriate title. 7. Adjust the class width using the following steps: A. Double click on a bar to pull up "Format Chart Area" B. Click the "Axis Options" dropdown menu and choose "Horizontal Axis" C. Click on the histogram on the left D. Select the dropdown menu "Axis Options" and change the bin width to 5.	1. Use the saved spreadsheet you created for Activity 3 of your data. 2. Highlight the column that has the life expectancies for the sample size of 30. 3. Select "Insert" in the top menu. 4. To get the histogram, click the icon that looks like a histogram (all bars are blue) and select the histogram picture on the left. 5. A histogram should pop up on your data editor. 6. Click on the box that says, "Chart Title" and create an appropriate title. 7. Adjust the class width using the following steps: A. Double click on a bar to pull up "Format Data Series" B. Click the histogram icon on the left C. Click on the "Auto" drop-down menu and choose "Bin width" D. Change the bin width to 5 and click enter.

Sketch your histogram in the space below.
Students' histogram will look like this one:

Frequency



Years

The distribution of life expectancies is unimodal and slightly skewed to the left but could be considered approximately symmetric based on the histogram.

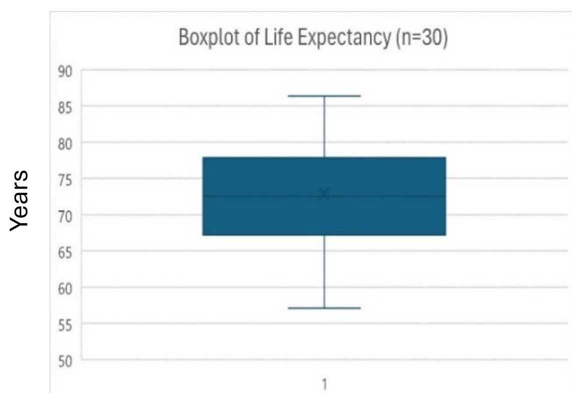
Note to teachers: The histograms will vary from one group to the next. Students will need to write in the units.

13. Use Excel to create a boxplot of the life expectancy data for the 30 countries you randomly selected in Activity 2. Does the boxplot show any outliers?
Note: Use the data from the spreadsheet you used for #12.

Steps for Creating a Boxplot:
1. Highlight the column that has the life expectancies for the sample size of 30.
2. Select "Insert" in the top menu.
3. To get the boxplot, click the icon that looks like a histogram and select the boxplot picture below the histogram.
4. A graph with one boxplot will pop up.
5. Click on the box that says, "Chart Title" and create an appropriate title.
6. Change the y-axis to better show the boxplots by following the steps below: A. Click on the y-axis numbers to pull up "Format Axis" B. Click on the icon that looks like a histogram C. Select the "Options" dropdown menu and change the bounds to 50 to 90. D. Click enter after each change.

Sketch your Boxplot in the space below.

Students' boxplot will look like this one:



The boxplot shows no outliers.

Note to teachers: The boxplot will vary from one group to the next. Students will need to write in the units by hand.

14. Argue the appropriateness of calculating the one sample t-distributions based on the sample sizes of n=5 and n=30?
    Your argument should primarily include comments that address characteristics needed to determine whether the second assumption is likely satisfied (e.g. symmetry and outliers).

    Note to teacher: The distributions could vary from one group to the next, so the answers may vary as well.

    Argument for sample size of 5:
    Based on the histogram and boxplot, the distribution of the data was approximately symmetric with no outliers. Technically this meets the assumptions, but when the confidence intervals are evaluated, they lack precision due to the small sample size.

    Argument for sample size of 30:
    Based on the histogram and boxplot, the distribution of the data was approximately symmetric with no outliers. Furthermore, the sample size is at least 30 allowing us to safely use the one sample mean t-confidence interval. With the larger sample size, the confidence intervals are more precise and useful.